Volume 5, Issue 2, February 2015 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

www.ijarcsse.com

Research Paper

Available online at: www.ijarcsse.com

Imagination, Detection and Mitigation of DDoS Attacks in Pdf File

Dipali Bhandwalkar, AmrutaBhoi, Karishma Salve, Chandrakala Bantanur, Prof. Mohan V. Pawar

Department of Computer Engineering

India

Abstract*— Distributed Denial of Service attacks are attempt to make a network resources become inaccessible, by interrupting services of host connected to network. Virus-affected computers are called as zombies. Botnet or robot network is an extensive group of zombie Computers. Botnet keeps the secret behind any file those are .mp3 and image file. These files deliver a request to certain intention selected by attacker and such zombie network cause incredible DDoS attack. The proposed system ensures that detection and mitigation of DDoS attacks in various file that comes under different module. We use two way detection in our system, first is signature based detection using snort, second is anomaly based detection using K-means clustering algorithm. The system ensures continuously learning new patterns and snort database updating after every new request, it also works for offline file based attacks. K-means is use to authorize new set of attack. K-means clustering algorithm gets combined with Snort to upgrade the detection of new malicious packet and decrease the redundant false alarm rate.

***Keywords*—DDoS Attack detection, Text based CAPTCHA, Snort, Anomaly Question Generation, Snort IDS, K-means clustering,Botnet, Zombie network.**

I. INTRODUCTION

Network security has become more important over the years observing the variation of attacks that the web services face every time. Various attacks like SQL injection, XSS injection, cookie capture calligraphy and DDoS come under this field. LOIC that is low orbit canon gun is an online tool which is used to carry out DDoS attacks. DDoS attacks aims to make the computer system resources inaccessible or it suspends the service of system. DDoS attacks are old but modern techniques used much advanced and insecure. In DDoS attacks, the attacker is capable to compromise a large number of computer hosts in the network and accomplish them for executing a corresponding attack. The power of DDoS attacks has varied with modern example of attack on github.com. DDoS termed as Distributed Denial of Service attack easily exhausted a communication from its victim and system resources is inaccessible to legitimate user. With the help of Botnet DDoS attack are done [2]. Robot network (Botnet) are collection of arbitrate machines that from network called as zombie network [3]. It constantly strike different website at a higher rate than common user can possible. DDoS attacks are illuminated in fig 1.

Online DDoS attacks can warn the services with the approach of cloud measuring or sharing of resources. Hence it becomes difficult to detect and reduce DDoS attacks. In a zombie network finding botnets can be mortal work as common client machines is uninformed of which machine is arbitrate and act as node to launch DDoS attacks. It is need to setup a system that will find and reduce DDoS attacks and at a time add arbitrate machines in exclude and declare each of them over a development of time. DDoS attacks have become so complicated that now-a-days attacker don’t select specific machine to arbitrate it slightly they upload a file as .mp3 or .png and attach botnet behind it. In this way will produce action as and when attacker has set date and time for DDoS attack. At that date and time all the hosts begin to DDoS a specific website. In this case attacker is totally unknown and common user come under infraction system. Existing papers shows only how to work in network packet to find DDoS.In this paper we propose a system which is based on anomaly based and signature based detection and improvement. This ensures that our system is continuously learning new patterns and snort database is getting updated after every new request. Here we also work for offline file based attacks. The system is divided into modules as:

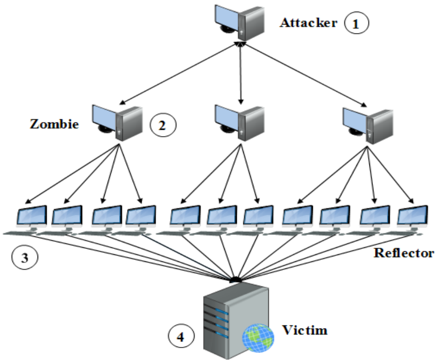


Fig. 1. Distributed Denial of Service attack.

Fig. 2.1. Snort: Snort is an open source NIDS that means any one is free to amend it.

Fig. 3.2. K-means clustering algorithm: Used to further forms of cluster to determine wicked packet or user.

Fig. 4.3.Turing test and question production module: It challenges user with text based questions. This question is to be answered appropriately in order to begin conversation with the server. The client that is incapable to answer the question appropriately is flagged as robot trying to initiate a DDoS attack.

II. RELATED WORK

DARPA (Defence Advanced Research Projects Agency) provided a network traffic datasets. Snort IDS is more efficient as it is open source that allows people to examine a package structure.[1]Shows only how to work on network packet to find DDoS attack but, here we also work for offline file based attacks. We proposed to use snort by considering objective that it is more efficient and execution level is moderately sophisticated than its challenger. [12] Snort established DoS detection system can be a real time effective and realistic execution that can counter changing DoS attack forms. [13] Shows that SNORT is Signature-Based IDS that uses combination of preprocessors and rules to examine network traffic. For detection of Signature based attack we use a snort IDS. Related work done in[4] Shows that they accomplish result in which large number of alerts have been minimized to 90%.[5]Shows that Anomaly based detection method detect those attack which are not known but provided more efficiency and those attacks in the database are detected by Signature based detection method.

K-means clustering algorithm is based on centroid points and can be said as centroid based clustering algorithm. [6] Shows the difference between K-means clustering algorithm and Fuzzy C-Means clustering algorithm. It shows that as compare to Fuzzy C-Means clustering algorithm, K-means clustering algorithm is more efficient according to its computational complexity. Hence we propose K-means clustering algorithm. For determining whether the client is Bot or normal user preceding work reflect use of image based CAPTCHA.[7]Adds falsification in form of noise in order to change image based CAPTCHA but this would be inflexible for normal user to distinguish and flag themselves as legitimate manipulators as it may necessitate various number of efforts. Hence this will not be useful. If the server keeps on giving image every time and attacker is assured to fire appeal at a rate much sophisticated than the server capacity this would ultimately DDoS itself as Image based DDoS attack. Because of that sending and receiving image much more complicated and it need greater bandwidth [8] than that of Text based CAPTCHA. Hence Text based CAPTCHA used in our propose system.

III. THE PROPOSED DDOS SYSTEM

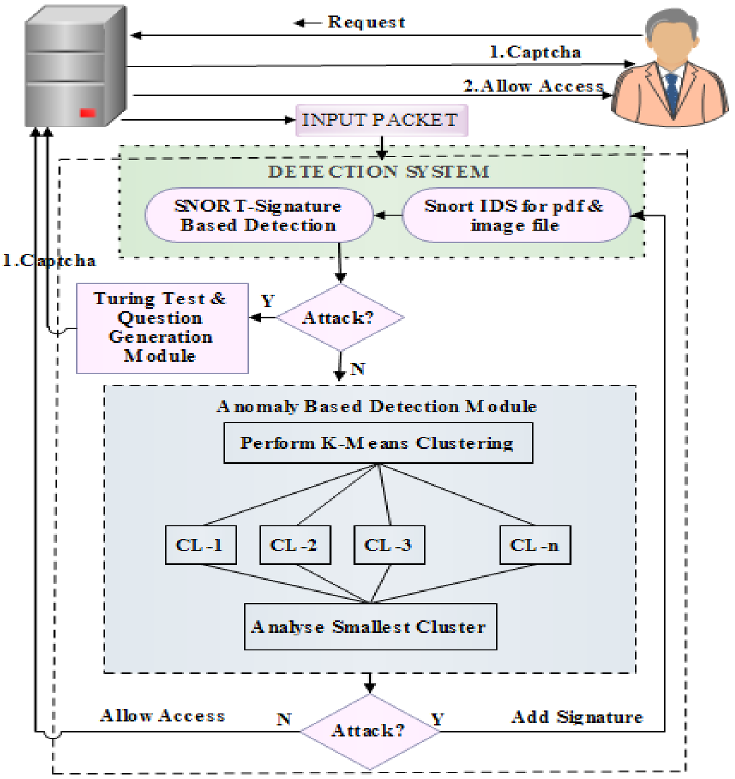

Request
1.Captcha
2.Allow Access
INPUT PACKET
DETECTION SYSTEM
SNORT-Signature Based Detection
Snort IDS for pdf & image file
1.Captcha
Turing Test & Question Generation Module
Y
Attack?
N
Anomaly Based Detection Module
Perform K-Means Clustering
CL-1
CL-2
CL-3
CL-n
Analyse Smallest Cluster
Allow Access
N
Attack?
Y
Add Signature

Fig. 5.Proposed DDoS Mitigation System.

Our methodology to detect and mitigate DDoS attacks is illustrated in Fig 2. Once the user has entreated for the web application; the server directs it to detection and mitigation system if its request arrives. If it is SYN flood attacks then it would not be efficient that the server answers the request. SYN flood attack is in which server assets its resources for the entire user and if user does not request to transmit on promotes communication that one a DDoS attack. The incoming package is first directed to Snort-Signature based detection module. Incoming package are compared with the signature of Snort in SNORT-Signature based detection module. If match is triggered then this is probably attacks to authorize it we send to Turing Test and Question generation module. In this module the user is reacted with a Text based CAPTCHA. Depending upon the response the system will adopt whether to authenticate the request else request is reject. The benefit of this is that we black list the pattern somewhat IP is not black listed this certifies higher accurateness than existing system. In other case in which Snort passes the input packet to next component which is Anomaly based detection. In this component we use K-means clustering algorithm to recognize the pattern of attack and form centroid points. This pattern if detected as an attack is further to Snort IDS for image and pdf file then add to Snort signature based detection this confirms compact Time computational complexity as if the similar packet reaches next time then it would be flagged as an attack in first module itself.

We propose a system that is based on Anomaly based detection and Signature based detection and mitigation. We use Snort in first module and K-means clustering algorithm in the second as the means to further form cluster to identify malicious packet or user. Turing test and Question generation module challenges user with the Text based question. This question is to be answered correctly in order to start communication with the server. The client that is unable to answer the question correctly is flagged as robot trying to launch a DDoS attacks. Signature based detection using Snort.

A. Snort Signature based detection

Snort is a vulnerable source NIDS that indicates any one is permitted to change and avail its source code.It is a signature-based IDS that usage a combination of rules and preprocessors to analyze network traffic. [9]Graceful influence network intrusion detection system is exhausted. This makes avail of Snort even more in expensive as we can amend its instruction set which establishes the most substantial fragment as packet are matched with these signature present in the Snort. Another advantage of using Snort is that it supports several outputs, saving alerts to records or database or generating a log for advanced exploration of the occasion triggered. In our proposed system we match incoming packet with the Snort rule set. A rule is defined as below:

<rule action><protocol><source ip><source port><direction><destination ip><destination port><rule options>

Example: alert tcp any any -> any 45

[Create an alert for any incoming packet send to port 45]

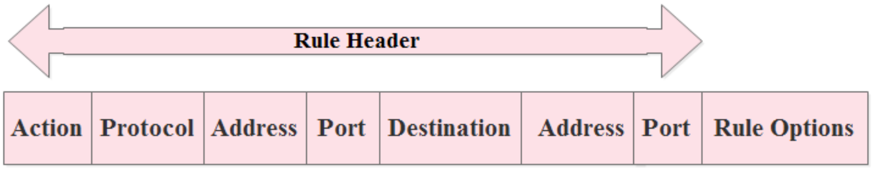


Fig. 6. Structure of a Snort rule.

Action: action part of the rule determines type of action taken when signature is matched. Actions can be an alert or logging of the event to the database.

Protocol: Protocol part determines the application of rule on packets for a particular protocol.

Address: Address part are be source and destination address.

Port: Port determines the source and destination ports of a packet on which rule are applied.

Direction: Direction part of rule determines which address and port number is used.

Components of Snort are illustrated below:

Packet Sniffer: It revenues packets from several network interfaces and formulate the packet to be preprocessed. That Packet will be sent to internal detection system.

Preprocessor: The Preprocessor accomplishes a change of preprocessing other than the standard packet decoding, before the data can be examined by Detection Device. These embrace IP fragment assembly, TCP stream assembly, packet legend standardization, etc.

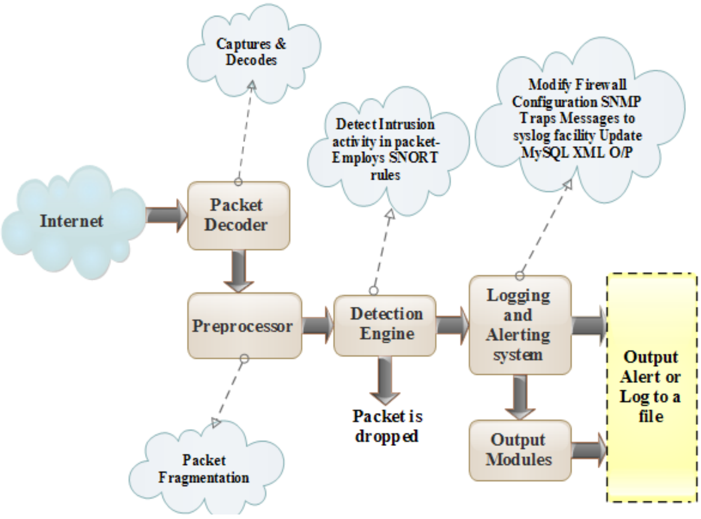

Captures & Decodes
Detect Intrusion activity in packet-Employs SNORT rules
Modify Firewall Configuration SNMP Traps Messages to syslog facility Update MySQL XML O/P
Internet
Packet Decoder
Preprocessor
Detection Engine
Logging and Alerting system
Output Alert or Log to a file
Packet is dropped
Output Modules
Packet Fragmentation

Fig. 7. Component of Snort.

Detection Engine: Here Snort rule set match with packet. If a match is activate then Snort sets an alert or logs the activity in its database. Accomplishment of Detection Engine is directly proportional to number of Snort rules and load on the system with respect to the incoming packets data send by preprocessor is examined through set of rule. If instruction matches the data in the packet then it is sent to alert system.

Alerting/Logging System: This last component logs or generates system alerts based on the action specified in the matched rules as well as the options given at the start of the system. Once the malicious packet is detected its activity is logged. These files are stored under /var/log/snort folder.

Database/Log files: Output generated by Logging and alerting system are control by this fragment. This can be thought as plug-ins that can do various operations depending upon output system that is required.

TABLE I WORKING OF SNORT

NAME	DESCRIPTION
Packet Sniffer	Preparation of packet for preprocessing
Preprocessor	Detect Anomalies
Detection Engine	Rule Application to the packet
Alerting & logging system	Generates alerts & create message logs
Log files	Generate file output

Snort is sustained on extent of stages and operating systems. Snorts are use under Linux and JAVA as interface language to detention and transmit packets to our detection and mitigation system. In our system to hold the signatures of malicious pattern of attacks we use Snort as a centralized database .The input is further delivered on to K-means clustering which is Anomaly based detection. If packet passes through Snort without any advice and if it warns as malicious packet then packet is passed to third module which is Turing Test and question generation module. In which client will be given with Text based CAPTCHA as exciting question to differentiate among normal client and an attacker.

B. Anomaly Based Detection

Clustering algorithm can recognize new attacks due to its advantages of unsupervised learning. Attacks with unknown signature pattern detected by clustering algorithm. Clustering algorithm assist in differentiating normal and intrusive traffic.

Two types of clustering algorithm

1. Coupling clustering algorithm.
2. Centroid clustering algorithm.

We use Centroid clustering algorithm over coupling clustering algorithm. For clustering CLUTO [10] toolkit is based on graph partitioning. CLUTO toolkit has proved that centroid clustering algorithm is efficient in handling large size of network traffic dataset using analysis. The complexity of coupling clustering algorithm is O(N^2), where n is the no of data instances. While that of Centroid clustering algorithm is O(KMN); Number of cluster is K;M is number of iteration and N is total no of data instances. K-Means clustering algorithm centroid based clustering algorithm is proposed to be utilized in the system. K-Means has the tendency of forming round clusters.

Suppose (a_1, a_2, …. a_n) are set of multi-dimensional real vector observations. K-Means clustering algorithms partitions these n observations into k sets such that k$\leq n$. Now to minimize the WCSS [within cluster sum of squares]

$$\arg\ min \sum_{\mathrm{i=1}}^{\mathrm{K}} \sum_{\mathrm{a_j \epsilon E_i}} \parallel \mathrm{a_i} - \mathrm{\mu_j} \parallel^2$$

Where μj is the mean of points in Ej.

Since we have to operate in 2 dimensional argument space, round clusters are desirable. For the clustering we will be using 2 arguments:

1. Total of alerts for each signatures
2. Quantity of Hosts causing alerts.

Euclidean distance metric be selected like it appropriate within case of centroid clustering algorithms which function within short- dimensional data spaces and as well since this metric is chosen for data investigation. The Euclidean distance flanked with points a and b is the distance of the fragment which joins these points.

If x={a_1 , a_2 , …. a_n} and y={ b_1 , b_2 , …. b_n }are two points in Euclidean space then distance from a to b is

□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ $\sqrt{\sum_{i=1}^{n} (a_i - b_j)^2}$

Davies-Bouldin index will be responsible for deciding the optimal amount of cluster.

If Ci is the cluster of vectors, Aj is an n-dimensional vector assigned to Cj , Ai is the centroid of Ci , Gi is the size of the cluster then Ei which is the measure of scatter is

$$E_i = \sqrt[2]{\frac{1}{G_i}\sum_{j=1}^{G_i} \mid A_j - M_i \mid^2}$$

If $R_{i,j}$ is the measure of separation between clusters C_i and C_j and $a_{k,I}$ is the kth element of M_i then

$$R_{i,j=} \parallel M_i - M_j \parallel = \sqrt[2]{\frac{1}{G_i}\sum_{k=1}^{n} \mid q_{k,i} - q_{ik,j} \mid^2}$$

Now, if Fi, is measure of how good the clustering scheme is

Then $F_{i,j} = \frac{E_i + E_j}{R_{i,j}}$

And $p_{i=max_{j:i\neq j}} F_{i,j}$

Then $PV = \frac{1}{N}\sum_{i=1}^{N} Pi$

Where PV is the Davies Bouldin Index.

The packet suffer from Snort will be input to this module for further analysis by K-means clustering algorithm packet will be subjected to analysis using the parameters mentioned above added to the Snort database. If the packet is found to be that of attacks then the signature retrieve from the cluster is passed on, else Server resources are accessed by client .Client is put to answer challenging question generated by Turing Test and question generation module once the signature is added to snort database. Client will be granted access to server resources only depending upon the authenticity of answer.

C. Turing Test and Question Generation Module

Turing test technique used in all website to discontinue the robots from attacking the website with spam comment these module intentions to challenge user with text based CAPTCHA [11]. A signature is created for each user that governs whether a user is suspicious or not[15].Design of a CAPTCHA is very complicated task, in case of image based CAPTCHA user has read the image and write the character in input box and submit it. If entered character are correct then image is posted otherwise rejected. Problem with this image based CAPTCHA are sending and receiving image is much more complicated as compared to text based CAPTCHA. So in our proposed system we use a text based CAPTCHA.

Various popular web based services utilize Image CAPTCHA like Google, Yahoo and several others. Disadvantage with conventional Text based string is that any robot would analysis the resource and then copy pastes the string and by-pass the authentication but we maintain a database in our proposed system which has collection of various queries for example:

What is Fifty– 5?

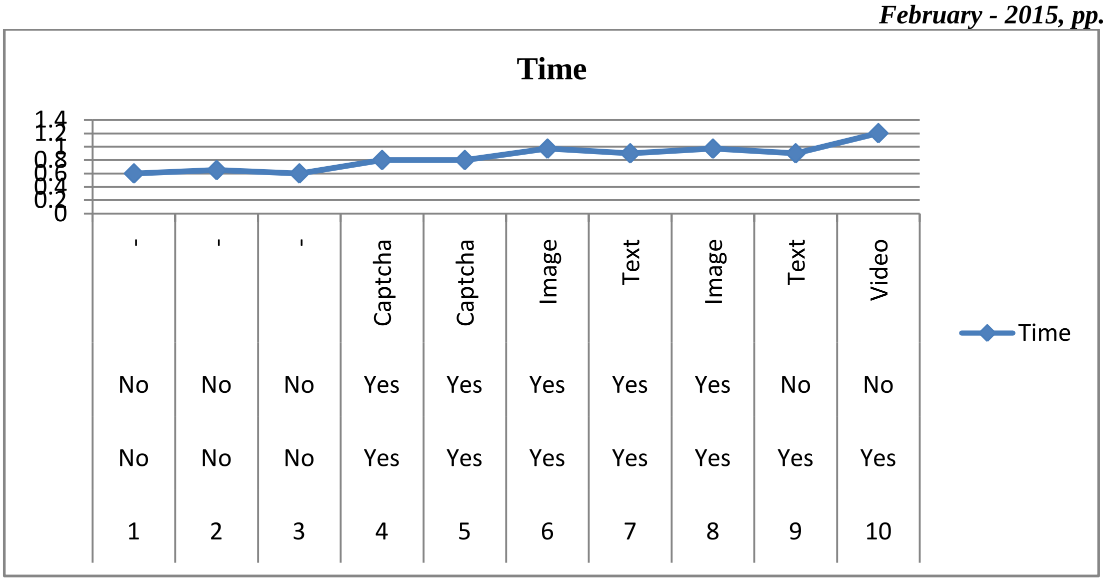
Which of 35, 100, and 10 is highest?

No alternative will be offer to customer and he/she has to respond. In example Fifty - 5. It cannot be any text but a number. Main intention of using “Fifty” in its place of ‘5’ is that yet if robot is trying to analysis the source code of the website; it will not be able to recognize text minus a number system. Our system would control robot from unauthorized access. To ensure full security Questions are not repeated. A further approach would be using Lexical Functional grammar to produce such sentences arbitrarily. This technique would be well-organized but it has huge drawback that it will increase the complexity of the system to manifold. Proposed system is to detecting DDoS attacks in pdf files and image files and not attacks itself. Use of Lexical Functional grammar would add over head on the server to execute redundant calculations. Hence we propose to use Text based CAPTCHA directly from the database.

This is one of the Question generation modules and the task of Turing Test another one is to provide server paperwork a web interface in which user is able to observe logs and reports, way actions as and when they occur. We would be setting a threshold value for reply of difficult questions as it cannot be infinite. This threshold would either be 3 or 5 which is as per recent standard. The server admin has the rights to change this threshold value if there is need to.

IV. RESULT

Authentication	Attack Exist	Detected	Type	Time
1	No	No	-	0.6
2	No	No	-	0.65
3	No	No	-	0.6
4	Yes	Yes	Captcha	0.8
5	Yes	Yes	Captcha	0.8
6	Yes	Yes	Image	0.97
7	Yes	Yes	Text	0.9
8	Yes	Yes	Image	0.97
9	Yes	No	Text	0.9
10	Yes	No	Video	1.2



V. CONCLUSIONS

In this paper we proposed imagination, detection and mitigation of DDoS attacks in pdf file. Our system consist of both methodologies in detecting a DDoS attacks that is Snort signature based and Anomaly based detection which makes our system more efficient with constant learning and updating Snort database. Overall flow of modules and their detail working are included in our system. Future work would be detecting DDoS attacks in html and xml files. Our intention is to develop a parser and integrate in our existing system in order to full proof server from intrinsic as well as extrinsic threats in any organization. This can be further extended to commonly used file formats such as .doc, .csv and even image file extension like .png, .gif. JavaScript code is binder behind these files. Whenever a click event appears the code embedded in these files becomes live and is executed on local machine. This may be a threat to server or other resource from within the organization.

ACKNOWLEDGMENT

We would like to express our gratitude towards Prof. H. A. Hingoliwala whose support and consideration has been an invaluable asset during the course of this project. First we would like to thank our guide Prof. M. V. Pawarfor providing us with their invaluable guidance throughout the course of this project. It would have been an almost impossible task to complete this project without his support, motivation, valuable suggestions and criticism.

REFERENCES

[1] A.Merani, N.Varghese, R.Deshmukh,“DDoS Detection and Mitigation using Snort, K-means clustering algorithm and Text based CAPTCHA” IEEE Internet Computing, International Conference on Convergence of Technology – 2014 978-1-4799-3759-2/14/$31.00©2014 IEEE 5

[2] G. Goth, “Fast-moving zombies: Botnets stay a step ahead of the fixes,” IEEE Internet Computing, vol. 11, pp. 7–9, 2007.

[3] P. Salvador, A. Nogueira, U. Franca, and R. Valadas, “Framework for zombie detection using neural networks,” in Internet Monitoring and Protection, 2009. ICIMP ’09.Fourth International Conference on, May 2009, pp. 14 – 20.

[4] A. Alharby and H. Imai, “IDS False Alarm Reduction Using Continuous and Discontinuous Patterns,” Applied Cryptography and Network Security, vol. 3531, 2005, pp. 423-442, doi: 10.1007/11496137_14.

[5] M. Ali Aydin *, A. HalimZaim, K. GökhanCeylan, “A hybrid intrusion detection system design for computer network security,” Computers and Electrical Engineering 35 (2009) 517–526.

[6] SoumiGhosh, Sanjay Kumar Dubey, “Comparative Analysis of K-Means and Fuzzy C-Means Algorithms, “International Journal of Advanced Computer Science and Applications, 35, Vol. 4, No. 4, (2013)

[7] J.Yan and A. S. El Ahmad, “Usability of captchas or usability issues in captcha design”, Proceedings of the 4th symposium on Usable privacy and security, 2008, pp. 44-52.

[8] R. Datta, J. Li, and J. Z. Wang, “Imagination: a robust image based captcha generation system”, Proceedings of the 13th annual ACM international conference on Multimedia.

[9] VivekVasishtha, Durgesh Kumar, “IDS Improved with K-Means Algorithms, Self-Organizing Map And Auto Class”, volume 2, issue 5, May 2012 ISSN: 227 128X. International Conference on Convergence of Technology – 2014 978-1-4799-3759-2/14/$31.00©2014 IEEE 5

[10] SHI ZHONG and TAGHI M. KHOSHGOFTAAR, “Clustering- Based Network Intrusion Detection,” Vol. 14, No. 2(2007) 169-187

[11] L. V. Ahn, M. Blum, N. J. Hopper, and J. Langford, “Captcha: using hard ai problems for security”, Proceedings of the 22nd International conference on Theory and applications of cryptographic techniques, 2003, pp. 294-311.

[12] Mothalla Saritha and Mukesh Chinta“Countering Varying DoS Attacks using Snort Rules”,International Journal of Advanced Research in Computer Science and Software Engineering,Volume 3, Issue 10, October 2013 ISSN: 2277 128X

[13] Ugur Akyazi and A. Sima Uyar, “Distributed Detection of DDoS Attacks During the Intermediate Phase Through Mobile Agents,” Computing and Informatics, Vol. 31, 2012, 759–778

[14] Byung-Jun Oh,Sang-Heon Shim and Young-Chon Kim“A Study on Recent Approaches in Handling DDoS Attacks”, Advanced Communications & Networks Lab, Division of Electronics & Information Engineering Chonbuk National University 561-756 Jeonju, Republic of Korea

[15] Naga Shalini Vadlamaniand Dr. Ju-Yeon Jo“A Survey on Detection and Defense of ApplicationLayer DDoS Attacks”,12-1-2013

[16] Neelam Paliwal, Ramesh Singh Rawat, Deepak Singh Rana,“Survey of Botnet Based DDoS Attack and Recent DDoS Incidents”International Journal of Advanced Research in Computer Science and Software Engineering,Volume 4, Issue 5, May 2014 ISSN: 2277 128X